

# Large-scale drivers and seasonal predictability of extreme wind speeds over the North Atlantic and Europe

Walz, Michael A.; Donat, Markus G.; Leckebusch, Gregor C.

DOI:  
[10.1029/2017JD027958](https://doi.org/10.1029/2017JD027958)

License:  
Creative Commons: Attribution (CC BY)

*Document Version*  
Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*  
Walz, MA, Donat, MG & Leckebusch, GC 2018, 'Large-scale drivers and seasonal predictability of extreme wind speeds over the North Atlantic and Europe', *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 20, pp. 11,518-11,535. <https://doi.org/10.1029/2017JD027958>

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**  
Checked for eligibility: 12/11/2018

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



## RESEARCH ARTICLE

10.1029/2017JD027958

## Key Points:

- Drivers and seasonal predictability of European winter storminess are explored using the ECMWF System 4 and a statistical model
- ECMWF System 4 does not reproduce the atmospheric variability patterns associated with enhanced storminess in the reanalysis
- Dynamical prediction skill is low, but the statistical model based on autumn conditions shows some predictability based on autumn MSLP conditions

## Supporting Information:

- Supporting Information S1

## Correspondence to:

M. A. Walz,  
maw526@bham.ac.uk

## Citation:

Walz, M. A., Donat, M. G., & Leckebusch, G. C. (2018). Large-scale drivers and seasonal predictability of extreme wind speeds over the North Atlantic and Europe. *Journal of Geophysical Research: Atmospheres*, 123. <https://doi.org/10.1029/2017JD027958>

Received 24 OCT 2017

Accepted 27 AUG 2018

Accepted article online 9 OCT 2018

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Large-Scale Drivers and Seasonal Predictability of Extreme Wind Speeds Over the North Atlantic and Europe

Michael A. Walz<sup>1,2</sup> , Markus G. Donat<sup>2,3</sup> , and Gregor C. Leckebusch<sup>1</sup>
<sup>1</sup>School of Geography, Earth and Environmental Science, University of Birmingham, Birmingham, UK, <sup>2</sup>ARC Centre of Excellence for Climate System Science and Climate Change Research Centre, University of New South Wales, Sydney, New South Wales, Australia, <sup>3</sup>Barcelona Supercomputing Center, Barcelona, Spain

**Abstract** As extreme wind speeds are responsible for large socioeconomic losses in the European domain, a skillful prediction would be of great benefit for disaster prevention as well as the actuarial community. Here we evaluate the patterns of atmospheric variability and the seasonal predictability of extreme wind speeds (e.g., >95th percentile) in the European domain in the dynamical seasonal forecast system European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 and compare to the predictability using a statistical prediction model. Further we compare the seasonal forecast system with ECMWF Re-Analysis (ERA)-Interim in order to advance the understanding of the large-scale conditions that generate extreme winds. The dominant mean sea level pressure patterns of atmospheric variability show distinct differences between reanalysis and System 4 as most patterns in System 4 are extended downstream in comparison to ERA-Interim. This dissimilar manifestation of the patterns across the two models leads to substantially different drivers associated with the generation of extreme winds: While the prominent pattern of the North Atlantic Oscillation could be identified as the main driver in the reanalysis, extreme winds in System 4 appear to be related to different large-scale atmospheric pressure patterns. Thus, our results suggest that System 4 does not seem to capture the potential predictability of extreme winds that exists in the real world. This circumstance is likely related to the unrealistic representation of the atmospheric patterns driving these extreme winds. Hence, our study points to potential improvements of dynamical prediction skill by improving the simulation of large-scale atmospheric variability.

## 1. Introduction and Motivation

Winter windstorms represent one of the most dangerous and loss-intensive natural hazards for the European region. According to the European Environmental Agency (2011), storms were the costliest natural hazards in Europe between the years 1998 and 2009, exhibiting an accumulated loss of more than €44 billion. Thus, it would be of outmost value to provide useful predictions on seasonal scales as it would enable decision makers to take measures in order to minimize potential losses and most importantly to avoid casualties.

The demand for these longer-term “weather forecasts” exceeding the common 10-day prediction period has generally increased considerably over the last decade. One of the reasons for that is certainly the desire to minimize casualties and loss due to extreme weather events especially with respect to climate change (e.g., Easterling et al., 2000; Lowe et al., 2016). Due to the atmosphere’s chaotic nature, however, it is generally impossible to predict single weather events, not to mention extreme events, deterministically on a time scale exceeding 5–7 days. The reason for that is the nonlinearity of the atmospheric system that amplifies minuscule deviations in initial conditions into large disturbances at the end of a forecast period (Lorenz, 1963). This behavior of the atmosphere is often referred to as the “butterfly effect.” There is, however, an intrinsic predictability in atmospheric variables on a longer time scale. This predictability is dependent on atmospheric and oceanic conditions that feature variability modes on longer time scales (e.g., the El Niño–Southern Oscillation or the Atlantic Multidecadal Oscillation; e.g., Knight et al., 2006). The Atlantic Multidecadal Oscillation is known to have an impact on the decadal variability of the North Atlantic storm track (Nissen et al., 2014) and also on its position (Woollings et al., 2012).

As a response to the rising demand of forecasts for intermediate time scales, the gap between day-to-day weather forecasts and long-term climate projections has been closed by seasonal predictions (Doblas-Reyes et al., 2013; Palmer et al., 2004). These seasonal predictions are based on free-running coupled atmosphere-ocean models that are usually initialized on the first day of each month. The initial conditions are

taken from the observed state of the atmosphere and the ocean on the given day. In most cases seasonal forecasts run for a period of 6–7 months. To account for the chaotic nature of the atmosphere, they are run as ensemble forecasts that often feature 20–50 ensemble members to enable probabilistic predictions.

As for every forecast product, there is always the question of forecast skill. Della-Marta et al. (2010) investigated the skill of the European Centre for Medium-Range Weather Forecasts (ECMWF) System 3 (Anderson et al., 2007) with regard to high wind speeds and found skill for the 95th percentile of wind speeds for the first lead month but none thereafter. Torralba et al. (2017) examined the skill of ECMWF System 4 (Molteni et al., 2011; System 4 hereinafter) regarding the potential wind energy yield (thus seasonal mean wind conditions) for two different regions featuring a substantial amount of wind farms (Canada and the North Sea). They showed that the seasonal mean wind predictions of System 4 have to be bias corrected in order to be useable for the end user (i.e., wind farms). The bias-corrected seasonal mean wind speeds, however, show skill especially in the tropics and also in areas relevant to wind energy production, for example, extratropics. MacLeod et al. (2018) also investigated the wind energy yield by performing statistical simulations based on subdaily (6-hourly) to monthly mean wind speeds taken from two reanalysis data sets. They concluded that only if there is predictable information in daily values will there be an information gain in using (sub)daily values compared to weekly to monthly means. As the predictability in seasonal forecast products is based on low-frequency variations of the climatic system, they are expected to have more skill on a weekly to monthly scale rather than a daily scale.

Extreme and mean winds in Europe are related to the North Atlantic Oscillation (NAO; e.g., Hurrell, 1995) as the leading mode of atmospheric variability for Europe (e.g., Pinto et al., 2009; Donat et al., 2010) especially for the winter months. Thus, a skillful prediction of the NAO could provide valuable information on potential storm-related impacts for upcoming winter seasons way ahead of common weather forecasts. The U.K. Met Office Global Seasonal Forecast System 5 (GloSea5; MacLachlan et al., 2015) has been shown to have significant skill for predicting the NAO for the European winter period (Scaife et al., 2014). Eade et al. (2014), however, argued that GloSea5 (and potentially other systems) underestimates the inherent predictability of the climate system (including the NAO) due to the noise in each of the single-ensemble members. They proposed the use of large ensembles in order to reduce noise and also a postprocessing method to adjust the variance of the ensemble prediction. Recently O'Reilly et al. (2017) investigated century-long seasonal hindcast simulations (uncoupled, System 4 based; see Weisheimer et al., 2017) and found a high variability in prediction skill of the NAO and the Pacific/North American (PNA) index. In particular, they identified a strong link between the lack of skill for the PNA and the sea surface temperature anomalies in the Pacific during the midtwentieth century in the seasonal hindcast experiments (for this particular case).

Moving from the potential predictors of winter windstorms toward the actual event, Bafort et al. (2018) investigated the capability of GloSea5 and System 4 to predict the frequency of occurrence of winter windstorms per season. They found significant positive correlations between the observed and forecasted number of windstorms on a grid cell level in large areas over the Atlantic Ocean for the years 1992–2011. However, they observed a drop in significance when investigating the skill for the longer 1983–2011 time span. Walz et al. (2018) in turn showed that a considerable amount of the interannual variability of these winter (December–January–February [DJF]) windstorm counts can be explained by large-scale Atlantic drivers (e.g., NAO, Scandinavian [SCA] pattern, or the East Atlantic [EA] pattern). Our study can be considered as a nexus between these two before-mentioned papers as a direct link between large-scale variability patterns and European winter extreme winds (in seasonal forecasts in particular) has not been shown to date. In order to investigate the predictability of extreme wind speeds, we are therefore trying to answer the question in how far the NAO (or other large-scale variability patterns) is associated with extreme wind speeds (associated with winter windstorms) in System 4. Thus, there are four questions we are trying to answer in this study:

- How well are winter large-scale mean sea level pressure (MSLP; e.g., NAO and SCA) patterns generally represented in the ECMWF System 4?
- What is the predictability of extreme wind speeds in seasonal forecasts? In other words, does the forecast contain more information than the climatology?
- Which large-scale drivers are responsible for extreme wind speeds in System 4 and in the reanalysis?
- How does a statistical model compare to the dynamical seasonal forecast regarding the predictability of extreme wind speeds to occur?

The next section of this paper features a description of the seasonal forecast system and reanalysis product used for this study. Section 3 introduces the methods used for this analysis. This includes the empirical orthogonal function (EOF) analysis of the MSLP data, the concept of statistical entropy as a skill measure, the quantification of the extreme wind speeds, and the description of the statistical model used to predict extreme wind speeds for the DJF period employing MSLP data from October/November. The results of this analysis will be presented in section 4 before concluding this paper with a summary and discussion of these results.

## 2. Data

The predictability of extreme wind speeds (using 6-hourly wind speed) and MSLP variability modes in the retrospective forecasts is investigated for the ECMWF Seasonal Prediction System 4 (System 4; Molteni et al., 2011). Six-hourly wind speeds have been used extensively to assess extreme winds, for example, associated with winter windstorms (e.g., Nissen et al., 2014, or Leckebusch et al., 2008). As the focus is on wind speeds associated with winter windstorms, boreal winters from 1982/1983 to 2013/2014 are used, which results in 32 DJF winter seasons. The reforecasts are initialized every 1 November, which corresponds to a lead time of 2–4 months when analyzing DJF. The atmospheric initial conditions are taken from the ECMWF Reanalysis (ERA)-Interim reanalysis (Dee et al., 2011). The ocean is initialized using the Ocean ReAnalysis System 4 (Molteni et al., 2011). Every reforecast comprises 51 ensemble members and is provided on a spectral resolution of T255, which is the same resolution used for ERA-Interim. The perturbed initial conditions are produced using singular vectors and an ensemble of ocean conditions of the ocean model Nucleus for European Modelling of the Ocean (NEMO) (Madec, 2008). The forecast system is based on the Integrated Forecasting System cycle 36r4 of the ECMWF. In the absence of high-quality, homogeneous wind observations for a number of locations across Europe, the predictability/skill of System 4 is assessed against ERA-Interim. We note that as both System 4 and ERA-Interim are based on similar model versions at the same resolution, there is only a minor bias in the simulated wind speeds, which should favor the detection of prediction skill. The wind fields for the North Atlantic/European domain (70°W–40°E, 30–70°N) are remapped to a  $5 \times 5^\circ$  grid for the predictability assessment to minimize the effects of small-scale spatial noise, in line with recommendations for decadal prediction verification (Goddard et al., 2013).

## 3. Methods

### 3.1. EOF Analysis of MSLP Data

EOFs are described by the eigenvectors of the covariance matrix of time series at different spatial points (e.g., Ambaum et al., 2001; Jolliffe, 1986). By definition, all eigenvectors are orthogonal to each other. EOFs represent an ideal tool to investigate the spatial and temporal variability of any given variable as each EOF explains the maximum temporal variance among all spatial fields. Usually, the EOFs are ranked by the amount of total variability that they explain. Thus, the first EOF explains the most temporal variability at any given point, the second EOF explains the second most variability, and so on. When using Northern Hemispheric (NH) MSLP data, the first EOF or eigenvector usually corresponds to the NAO as it is responsible for most of the variability in the NH (e.g., Ulbrich & Christoph, 1999, or Woollings et al., 2015). The second mode is usually associated with the PNA (e.g., Ambaum et al., 2001; naming of the EOF modes in the reanalysis in accordance with National Oceanic and Atmospheric Administration, <http://www.cpc.ncep.noaa.gov/data/tele-doc/telecontents.shtml>), which is the leading pattern of interannual variability of MSLP for the Pacific region.

For this study the EOF analysis was carried out for the standardized (by dividing by the standard deviation) DJF MSLP anomalies of the entire NH (20–90°N) for the ECMWF System 4 and ERA-Interim to compare their spatial patterns. Due to the constraining (orthogonal) nature of the EOF analysis, it turns out that the patterns of the two models look very different (see section 4.1. for details). In order to obtain comparable EOF time series for further analysis (especially with regard to drivers of extreme winds), we projected the MSLP anomalies of System 4 onto the eigenvector loadings of ERA-Interim. That way, we are aiming for a fairer comparison between the different driver time series of Season 4 and ERA-Interim. For completeness the results employing the System 4 eigenvectors are shown in the supporting information.

For the process of developing the statistical prediction model the EOF analysis was also done for the same region for MSLP data from 15 October to 15 November for every year for ERA-Interim. In order to make use of the large ensemble and to reduce the noise of each ensemble member, all 51 members are pooled

together to calculate joint EOFs. If the EOF was calculated for each of the members individually, the variability of the ensemble mean would be too low (Figure 8). The first nine EOFs were calculated for both System 4 and ERA-Interim. The loadings or principal components of the EOF analysis represent the time series of the respective EOF pattern. For comparison reasons the NAO index was also calculated as a normalized pressure anomaly difference between the grid boxes closest to Lisbon and Reykjavik from the ensemble mean of the hindcasts in System 4.

### 3.2. A Normalized Sum as a Measure for Extreme Wind Speeds

As extreme wind speeds (represented by percentiles) tend to be of fairly erratic nature with large year-to-year variability, the percentile alone does not show a high predictability on an interannual time scale (see Figure S1 in the supporting information). For that reason we chose to estimate the predictability of System 4 extreme wind speeds by incorporating the entire upper tail of each local wind speed distribution in order to get a more robust target variable. This is implemented by summing up all wind speed values exceeding a certain percentile threshold and normalizing this sum with the total amount of available time steps per season  $n$  (see equation (1)). The resulting value  $I_{k,y,9X}$  represents an integrated measure over the amount of percentile exceedances as well as the magnitude of exceedances on an interannual time scale. Accordingly, a large value could represent either a season with many smaller exceedances or a season with few large exceedances of the respective percentile.

$$I_{k,y,9X} = \frac{\sum_1^n v_{n,k}}{n}, \forall v_{n,k} > v_{9X,k} \quad (1)$$

We chose the 90th, 95th, and 98th percentiles as thresholds for this study as these percentiles have been used previously with regard to extreme wind speeds (e.g., Leckebusch et al., 2008, or Della-Marta et al., 2010). The percentile is calculated as a climatological percentile from the mutually available data (1982–2014) for forecasts and observations separately in order to avoid reducing predictability by a potential bias.  $I_{k,y,9X}$  is calculated for every grid cell  $k$  for every DJF season  $y$  between 1982/1983 and 2013/2014 for all 51 ensemble members and compared to ERA-Interim for all three percentiles. In order to assess the predictability in the European region,  $I_{k,y,9X}$  is averaged for time series comparison for the box shown in Figure 6.

### 3.3. Statistical Entropy and Predictive Power

The concept of statistical entropy has its roots in information theory and was first introduced by Shannon (1948, 1951) who defined it as a measure to estimate the information content of a message. Schneider and Griffies (1999) and DelSole (2004) took up on this idea and developed a conceptual framework of how to use statistical entropy as a measure of predictability. They argue that a forecast is best described by its entire distribution and that valuable information is lost when the distribution is reduced to mere moments (e.g., root-mean-square error (RMSE)). As this section is merely adequate to do justice to the complex field of predictability and information theory, we refer the reader to the excellent original papers by Schneider and Griffies (1999), DelSole (2004), and also Tang et al. (2007) for a detailed disquisition of the topic. This section shall provide a short and more qualitative overview over the concepts that will be used in this paper.

The Shannon (statistical) entropy  $S_X$  of a random variable  $X$  is defined as follows (Shannon, 1948):

$$S_X = -\int dx p_X(x) \log p_X(x) \quad (2)$$

$S_X$  is a measure of uncertainty related to the random variable  $X$ . If all realizations of the random variable  $X$  are equally likely,  $S_X$  is maximal as the outcome of the realization is hardest to predict. An example would be the rolling of a fair dice as every number has a probability of 1/6. Thus, every roll of the dice conveys the maximum amount of information. If the dice was biased, say, the number 3 was on four of the six sides, there would be less uncertainty as the number 3 can be expected to appear with a probability of 2/3. This reduction in uncertainty also reduces  $S_X$ . The extreme-case scenario would be a dice with the same number on every side. The uncertainty would be zero and so would be the Shannon entropy as the rolling of the dice does not convey any new information. This concept can be applied to ensemble forecasts: The uncertainty within an ensemble forecast depends on the spread of the ensemble members and would be maximal if every member would predict a completely different state of the atmosphere. The probability of every predicted atmospheric state would be 1/51 in case for System 4. If all ensemble members of the forecast ended up

in the exact same state, however, there would be no uncertainty at all in the forecast, and thus, Shannon's entropy would be zero. As follows, a robust forecast will produce smaller values of the statistical entropy as it is more constrained. However, the actual (correlation) skill of the forecast could still be small in the case that all ensemble members drift to an incorrect state. This would be considered an overconfident forecast.

A measure of predictability that makes use of these considerations is given by the predictive information (PI) as defined by Schneider and Griffies (1999) and adapted by Tang et al. (2007):

$$PI = S_X - S_E \quad (3)$$

$S_X$  describes the uncertainty (statistical entropy) associated with the climatological distribution, thus if no forecast was available (prior uncertainty). The variable  $S_E$  represents the statistical entropy after the forecast (and observation) has become available. It quantifies the uncertainty in the ensemble prediction (posterior uncertainty). Ideally, the posterior uncertainty is smaller than the prior uncertainty ( $S_E < S_X$ ) due to useful information in the ensemble forecast. Hence, a large PI value suggests a more reliable forecast. Practically, however, the uncertainty in ensembles can be larger (i.e.,  $PI < 0$ ) than the uncertainty in the climatology, especially for an erratic variable such as extreme wind. Grid cells with negative values are considered to feature no predictability/additional information to the climatology. In other words, the climatology provides more information (is "sharper") than the forecast. Another measure defined by Schneider and Griffies (1999) and adapted by Tang et al. (2007) is represented by the predictive power (PP):

$$PP = 1 - \exp(-PI) \quad (4)$$

Equivalently to the interpretation of the PI, large values of the PP suggest useful information in the seasonal ensemble forecast, making it superior to the climatology. Evidently, PP becomes maximal ( $PP = 1$ ) if the prior uncertainty is infinite and the posterior uncertainty vanishes. As the PP exhibits a proper limiting behavior ( $0 < PP < 1$ ), it represents a descriptive quantification of the predictability. For this paper the PP will be calculated for the previously defined variable  $I_{k,y,9X}$  for each winter individually and also as a time mean that refers to the average predictability of extreme wind speeds for every  $5 \times 5$  grid cell in the studied domain.

### 3.4. Akaike Information Criterion Selection of Large-Scale Drivers

In order to understand what drives the interannual variability of extreme winds in System 4 as well as in ERA-Interim, we examined which of the nine large-scale MSLP variability patterns (EOFs) are most significant for extreme winds in the North Atlantic and European regions. The selection criterion is based on the AIC (an information theoretic criterion, occasionally also referred to as Akaike information criterion; Akaike, 1974) stepwise approach. Similar to the statistical entropy, the AIC is based on information theory (Jaynes, 1957). The AIC is an estimate of how much information is lost by using a statistical model instead of the actual physical relation. Thus, the AIC can be used as a tool for model selection if different models can be compared to each other. The essential part of the selection process is the trade-off between the goodness of fit and the complexity of the model, as the number of parameters to be estimated ( $k$ ) as well as the maximum of the likelihood function  $L$  are part of the AIC score:

$$AIC = 2k - 2 \ln L \quad (5)$$

The winning model will exhibit the smallest AIC value. Different from other model selection criteria, for example, the  $F$  test, the AIC score does not provide any information about the absolute quality of the model, however. Thus, if all potential models provide a poor fit, there is still a winning model albeit it being of poor quality.

To account for this drawback, we developed a two-step algorithm to identify the main drivers of extreme wind speeds for the studied domain:

1. By applying the AIC, we determined the best model for each grid cell checking combinations of all nine computed leading EOF time series as potential predictors and  $I_{k,y,9X}$  as the predictand. As  $I_{k,y,9X}$  is sufficiently normally distributed, a multilinear regression model is a natural candidate. The simplest model would be a constant straight line; the most complex model would include all nine predictors. The AIC stepwise algorithm in *R* (MASS package) examines every possible combination of the nine predictors and estimates the model with the lowest AIC score, thereby providing the "best" combination of drivers for every grid cell.



2. To determine the driver explaining the most variability, the predictors of the winning AIC model for each  $5 \times 5$  grid cell are tested for significance using a  $t$  test. Any predictor not significant at the 95% significance level is discarded from the AIC model. For each grid cell, the predictor associated with the largest regression coefficient is then considered to be the most important (winning) large-scale driver for the respective grid cell. Due to the  $t$  test constraint, there can be grid cells where no significant driver can be determined. This two-step algorithm is applied to both System 4 and ERA-Interim so that the internal drivers explaining the most variability of extreme winds can be compared between the seasonal forecast and reanalysis. This could help to understand why the potential predictability in seasonal forecasts is sometimes lower than in the real world as discussed in Eade et al. (2014), as we can examine what drives these extreme wind speeds (model internally). In case the extreme wind speeds are caused by different drivers in System 4 compared to ERA-Interim a potential lack of skill in System 4 for extreme wind speeds could be associated with the disparate-generating drivers.

In addition to the presentation of the winning driver for every grid cell, we provide correlation maps of selected drivers to illustrate the different driving mechanisms between the seasonal forecast and the reanalysis.

### 3.5. A Multilinear Regression Model to Statistically Predict Extreme Wind Speeds

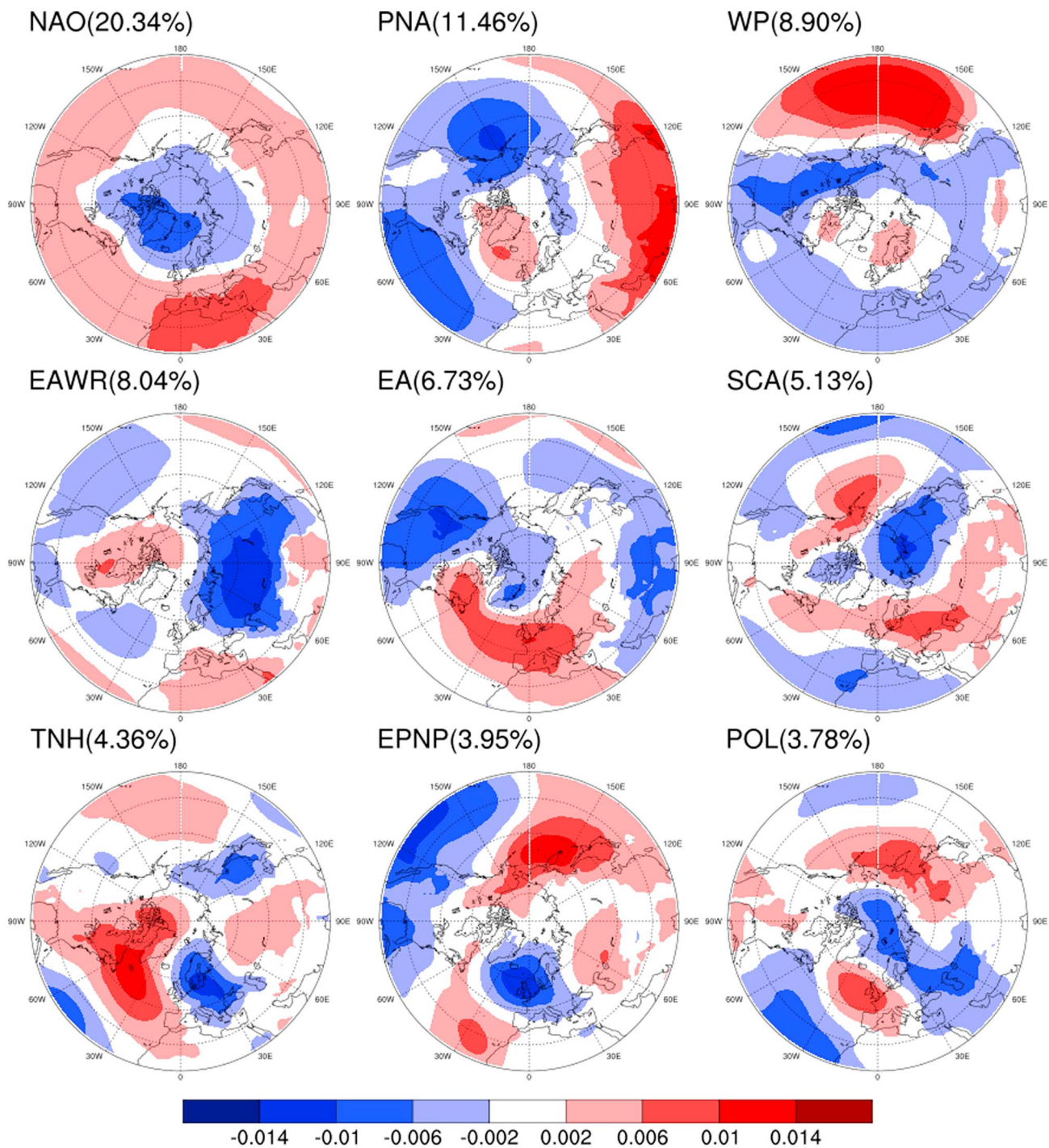
The statistical model is trained to predict the  $I_{k,y,9x}$  value for the upcoming DJF season based on the EOFs computed from MSLP ERA-Interim fields entailing data from 15 October to 15 November of every year. This period is centered on the initialization date of System 4 and thus imitates the lead time of 2–4 months of the dynamical seasonal forecast model. As we include data until 15 November, we do include more information in the statistical model than what would be available to System 4; however, we balance that by including data all the way back to 15 October. Therefore, we consider this as a fair base of comparison between the skill of the statistical and dynamical models. The procedure of training the model is similar to the identification of the large-scale drivers for the extreme wind speeds. The nine leading EOF patterns for the October–November period are used as a pool of predictors to predict  $I_{k,y,9x}$  for the coming DJF season. The model selection is equivalent to the algorithm described in the previous section: The selected predictors of the winning AIC model are tested for significance using a  $t$  test. As  $I_{k,y,9x}$  is sufficiently normally distributed, the natural choice for the model is a multilinear regression model again. To account for a potential overfitting of the model, we performed a twofold cross-validation of the winning model; thus, each year is used as a training set and as a testing set once. In case no significant driver can be determined for a grid cell, there is, of course, no statistical model for that respective grid cell.

The skill of the statistical model and the seasonal forecast model is examined by a correlation with the  $I_{k,y,9x}$  value for all available grid cells of the ERA-Interim reanalysis data. Particularly, the skill for Central and Northern Europe (as defined by the rectangular box in Figure 6) is analyzed.

## 4. Results

### 4.1. Differences in EOF Patterns Between System 4 and ERA-Interim

The nine leading EOF modes of DJF MSLP data are presented in Figures 1 and 2 for ERA-Interim and ECMWF System 4, respectively. Clearly, the leading EOF for both the dynamical prediction model and the reanalysis represents the expected NAO pattern featuring the Icelandic Low and the Azores High. The location of this dipole, however, appears different for the two data sets. It seems that the centers of these two pressure patterns are more downstream in System 4 compared to ERA-Interim. As a result, the pressure gradient over Europe appears smaller in System 4. The second EOF (PNA) also appears shifted downstream in System 4, reaching all the way into the European continent. Generally, it seems that there appears to be a model bias that “smears” the variability patterns in a downstream direction. Another example for that is the third EOF (Western Pacific [WP] pattern), which also seems to explain a considerable amount of variability in the European region in System 4 although its center of action should lie in the Pacific. In System 4 it appears almost as a mirrored NAO with high-pressure anomalies over Greenland and low-pressure anomalies over the Central Atlantic, inconsistent with the reanalysis-based EOF patterns. The sixth (SCA) EOF in comparison exhibits a very similar pattern for both ERA-Interim and System 4 with a blocking situation over Central Europe. The eighth (East Pacific/North Pacific) EOF with its dipole over the Pacific and a strong low-pressure

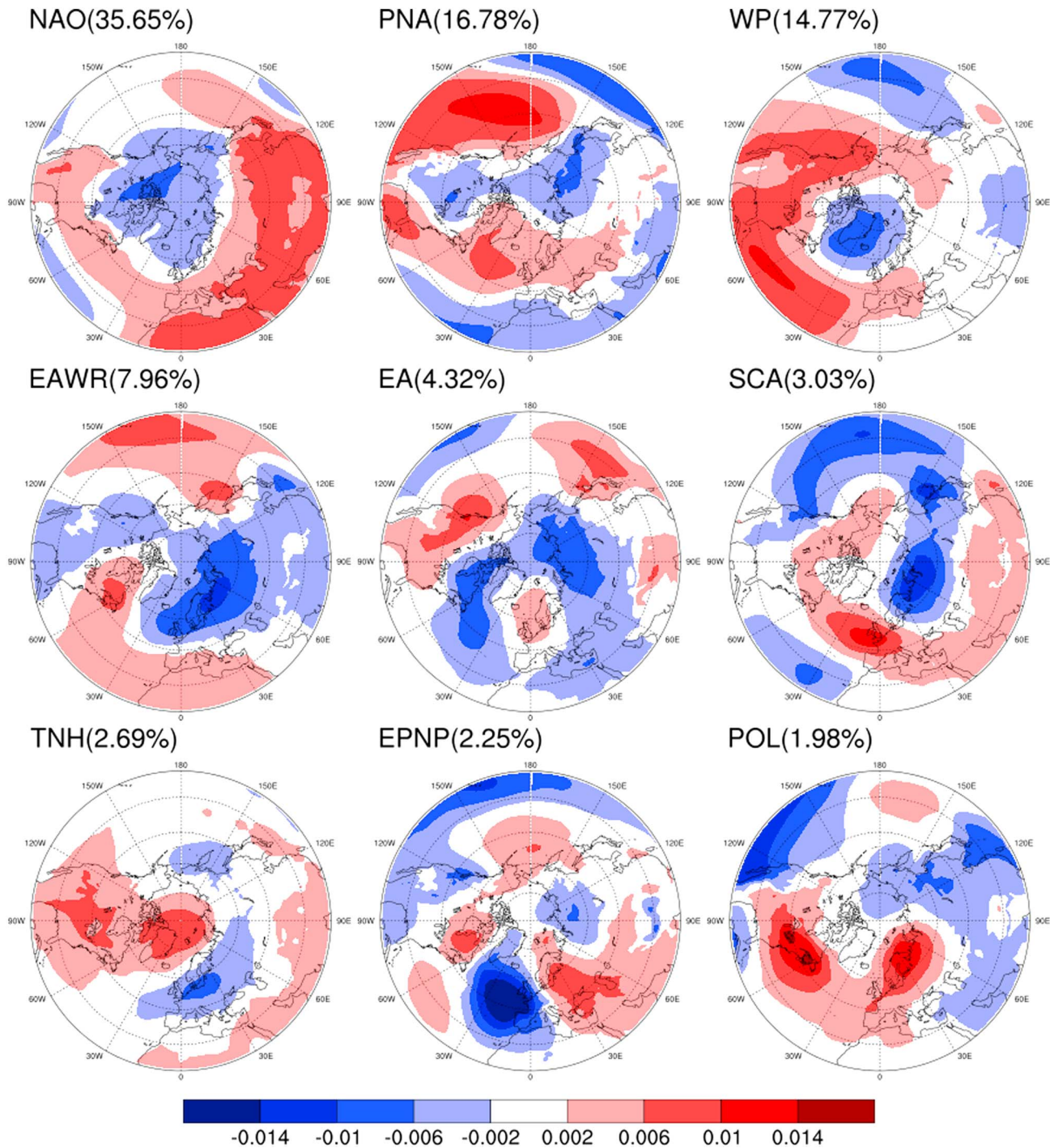


**Figure 1.** The nine leading EOF modes of MSLP during DJF in ERA-Interim. The explained variance of every mode is given after the conventional name of the EOF mode. EOF = empirical orthogonal function; MSLP = mean sea level pressure; DJF = December–January–February; ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; NAO = North Atlantic Oscillation; PNA = Pacific/North America; WP = West Pacific; EAWR = East Atlantic/West Russia; EA = East Atlantic; SCA = Scandinavia; TNH = Tropical/North Hemisphere; EPNP = East Pacific/North Pacific; POL = Polar.

feature over the British Isles is also comparable between the two data sets. Other EOF patterns, however, look completely different as, for example, the fifth (EA) and seventh (Tropical/Northern Hemisphere) EOFs.

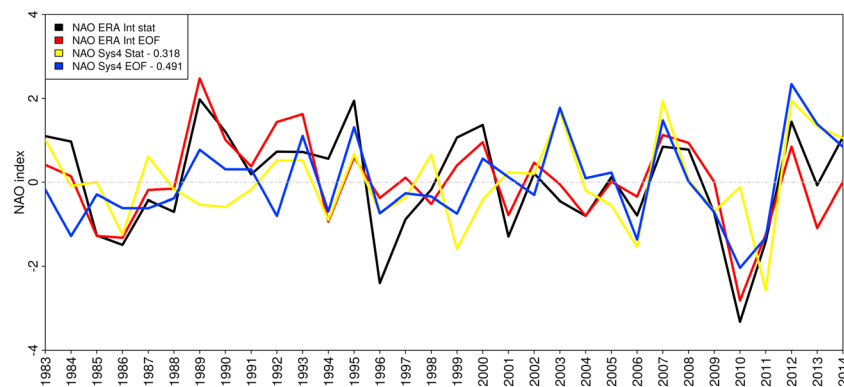
The explained variance by the different modes differs considerably between System 4 and ERA-Interim. Whereas the NAO accounts for around 20% of the interannual MSLP variability in ERA-Interim, it accounts





**Figure 2.** The nine leading EOF modes of MSLP during DJF in ECMWF System 4. The explained variance is given after the name each EOF mode would have if the NOAA nomenclature was used. EOF = empirical orthogonal function; MSLP = mean sea level pressure; DJF = December–January–February; ECMWF = European Center for Medium-Range Weather Forecasts; NOAA = National Oceanic and Atmospheric Administration; NAO = North Atlantic Oscillation; PNA = Pacific/North America; WP = West Pacific; EAWR = East Atlantic/West Russia; EA = East Atlantic; SCA = Scandinavia; TNH = Tropical/North Hemisphere; EPNP = East Pacific/North Pacific; POL = =Polar.

for about 35% in System 4. More generally, the explained variance (added up over all nine modes) for the nine leading EOFs in System 4 (~87%) seems to be considerably larger than for ERA-Interim (~72%). There is a more abrupt change in explained variance in System 4 between the fourth (East Atlantic/West Russia) and fifth (EA) EOFs, also. The decline of variance across the nine EOFs in ERA-Interim is smoother compared to that.



**Figure 3.** Time series for the EOF-based and the grid-box-based NAO index for System 4 and ERA-Interim. The correlation between the respective indices in ERA-Interim and System 4 is given in the legend. The years on the x-axis refer to the years of the January and February of a particular DJF season; thus, 1990 refers to the winter 1989/1990, and so on. EOF = empirical orthogonal function; NAO = North Atlantic Oscillation; ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; DJF = December–January–February.

As these EOF patterns manifest so differently in the two data sets, we decided only to use the EOF eigenvectors of ERA-Interim for calculating the time series and thus for identifying the drivers of extreme wind speeds as described in section 3.1.

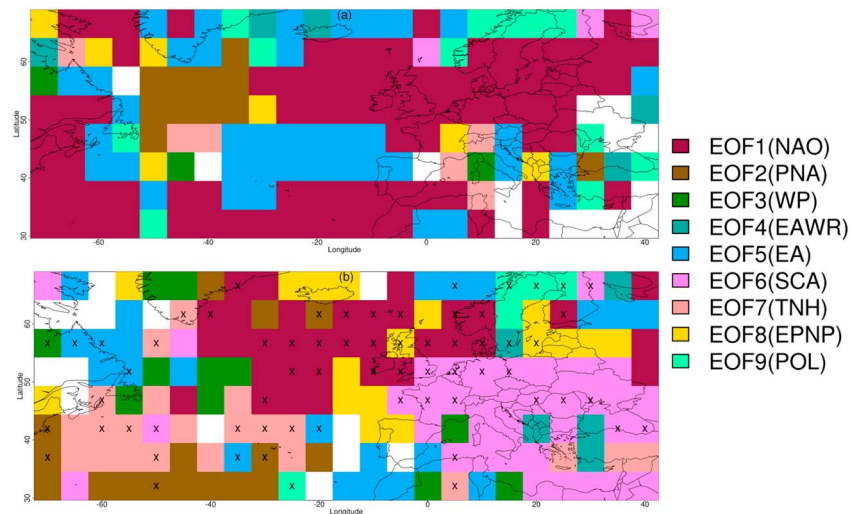
Figure 3 depicts the NAO time series of the EOF-derived indices as well as the grid-box-based indices. Note that the System 4 EOF-based index uses ERA-Interim loading patterns and System 4 MSLP anomalies. The correlation between the two respective EOF-based NAO time series is 0.49, whereas the correlation between the grid-box-based time series is only 0.32. Evidently, the EOF-based and grid-box-based NAO indices for System 4 appear to differ from each other considerably for certain years (e.g., 1984, 1999, or 2010) as well. The correlation between these two is 0.69, whereas the correlation between the two ERA-Interim indices is 0.80. The reason for the lower correlation of the System 4 indices could be due to the definition of the grid-box-based index using Lisbon and Reykjavik as reference locations. This could be similar to the findings of Ulbrich and Christoph (1999) who found that the center of actions of the NAO dipole shifts downstream for some general circulation model future climate projections.

#### 4.2. Drivers for Extreme Wind Speeds in ERA-Interim and System 4

The MSLP EOF pattern featuring the largest (absolute) regression coefficient for every grid cell is presented in Figure 4. Each color plotted on the map represents one of the nine leading EOF MSLP patterns using the nomenclature of the National Oceanic and Atmospheric Administration.

All predictors shown represent the “winning” coefficient for each grid cell, explaining most of the interannual variability of extreme wind speeds. Figure 4 depicts the drivers for  $I_{k,y,95i}$ ; however, the maps for  $I_{k,y,90}$  and  $I_{k,y,98}$  look very similar. As expected from numerous studies (Donat et al., 2010; Pinto et al., 2009), the NAO explains the largest part of the variability of high wind speeds for the ERA-Interim reanalysis as it is known to be a key driver for North Atlantic storminess. The areas where NAO has the strongest effects are split by the EA, which has also been identified as a variability pattern with a significant impact on European storminess (Walz et al., 2018, or Mailier et al., 2006). The extreme wind speeds in the Scandinavian region appear to be driven by the Polar (POL) pattern and the SCA (Walz et al., 2018).

The predictors for the anomalies of System 4 projected on the ERA-Interim loading patterns, however, draw a somewhat different picture: While the northern part of the NAO dipole looks similar to the one observed for ERA-Interim, the southern part associated with the Azores High does not seem to be a driver of the interannual variability at all. Instead, there is a mix of various drivers that seem to be responsible for extreme wind speeds for that area such as the EA or Tropical/Northern Hemisphere pattern. The absence of the southern part of the NAO as a driver might be due to the downstream displacement of the high pressure associated with the Azores High in the leading EOF pattern (Figure 2). The main driver for extreme wind speeds in

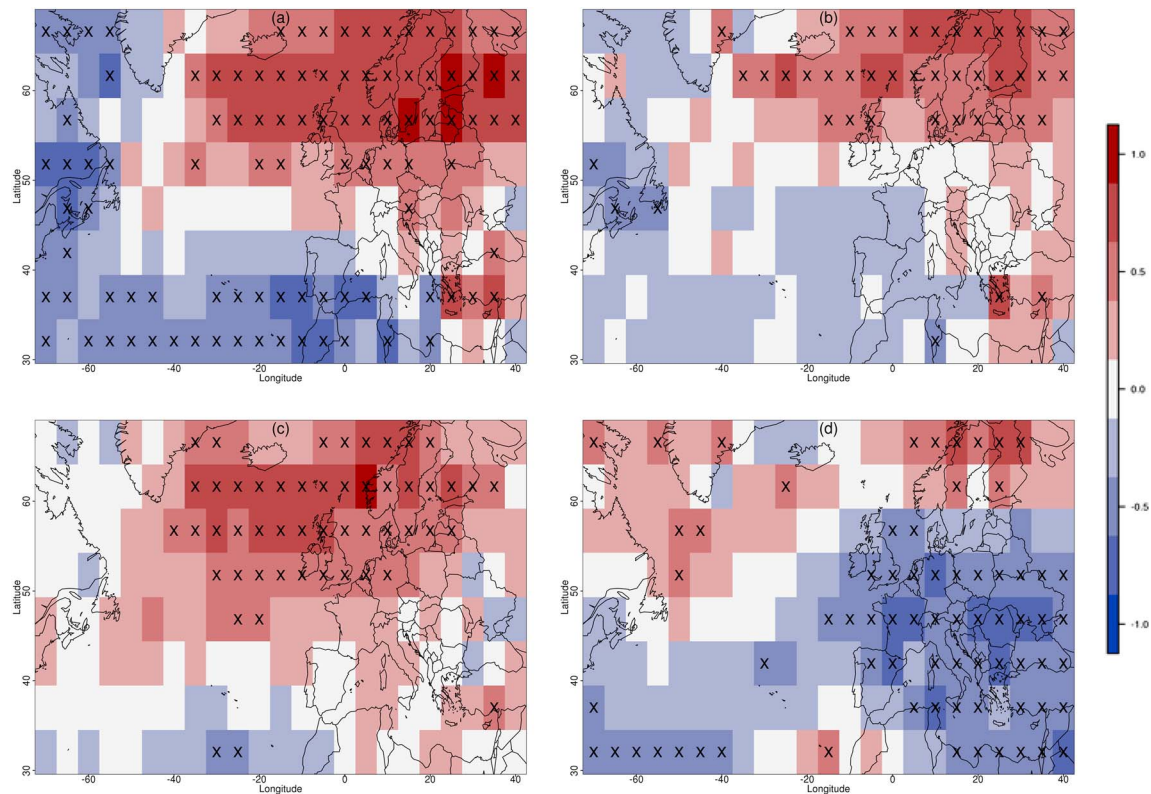


**Figure 4.** MSLP variability drivers associated with the largest absolute regression coefficient for every grid cell for explaining the interannual variability of  $I_{k,y,95}$ . If no significant driver could be determined, the grid cell remains white. (a) Winning drivers for ERA-Interim, and (b) winning drivers for System 4. The black crosses denote grid cells in which the winning driver of ERA-Interim (a) was at least among the top 3 for System 4. The nomenclature for the EOFs is according to the NOAA standards for the NCEP reanalysis. MSLP = mean sea level pressure; ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; EOF = empirical orthogonal function; NOAA = National Oceanic and Atmospheric Administration; NCEP = National Center for Environmental Prediction; NAO = North Atlantic Oscillation; PNA = Pacific/North America; WP = West Pacific; EAWR = East Atlantic/West Russia; EA = East Atlantic; SCA = Scandinavia; TNH = Tropical/North Hemisphere; EPNP = East Pacific/North Pacific; POL = =Polar.

Central Europe for System 4 appears to be the SCA blocking pattern, whereas it barely appears in ERA-Interim. The black crosses in Figure 4b denote grid cells in which the winning driver of ERA-Interim is at least among the top three drivers of System 4. This implies that the NAO might not be the most important driver for extreme wind speeds for the Central European region; however, it still explains a considerable amount of variability in System 4 (see Figure 5). The differences in drivers between the two plots (Figures 4a and 4b) can be explained by the different manifestations of the large-scale MSLP patterns. When interpreting these results, it has to be kept in mind, however, that the EOFs are constrained by their mutual orthogonality. This means that if a data set is a linear superposition of two patterns that are not orthogonal, the EOF analysis will not yield these patterns. Furthermore, the resulting EOF patterns can depend on the spatial domain (see Ambaum et al., 2001).

The differences between the identified drivers for the reanalysis and the seasonal forecast become more obvious when comparing the correlation maps for the NAO and SCA variability patterns with extreme wind speeds individually. The correlation between  $I_{k,y,95}$  and the NAO and the SCA for both ERA-Interim and System 4 are presented in Figure 5. Correlations significant at the 95% significance level are marked with a black cross. The correlation pattern of the NAO for ERA-Interim exhibits the prominent dipole pattern over the North Atlantic featuring positive correlations between the NAO and extreme wind speeds over Northern Europe. Negative correlations are present over Southern Europe and over the Atlantic between 30°N and 40°N. The correlation patterns look considerably different for ERA-Interim and System 4. Whereas the northern part of the correlation for the NAO looks similar in both data sets, the southern part barely shows any significant negative correlation in System 4. The difference in correlations for the SCA pattern is even more striking. Basically, the entire European continent features a negative correlation between the sixth EOF (SCA) time series and the interannual variability of extreme wind speeds in winter in System 4, whereas the main region of correlations in ERA-Interim appears to be (compliant with its name) the Scandinavian region. Also, the correlation pattern features different arithmetic signs. As the correlation between the SCA and extreme wind speeds in System 4 over Central Europe is fairly strong ( $<-0.5$ ), it explains why the SCA appears as the most important driver of the variability as seen in Figure 4.





**Figure 5.** Correlation maps between EOF1 (NAO) and  $I_{k,y,95}$  time series for (a) ERA-Interim and (c) System 4 and between EOF6 (SCA) and  $I_{k,y,95}$  time series for (b) ERA-Interim and (d) System 4. All significant correlations at the 5% confidence level are marked with a black cross. EOF = empirical orthogonal function; NAO = North Atlantic Oscillation; ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; SCA = Scandinavia.

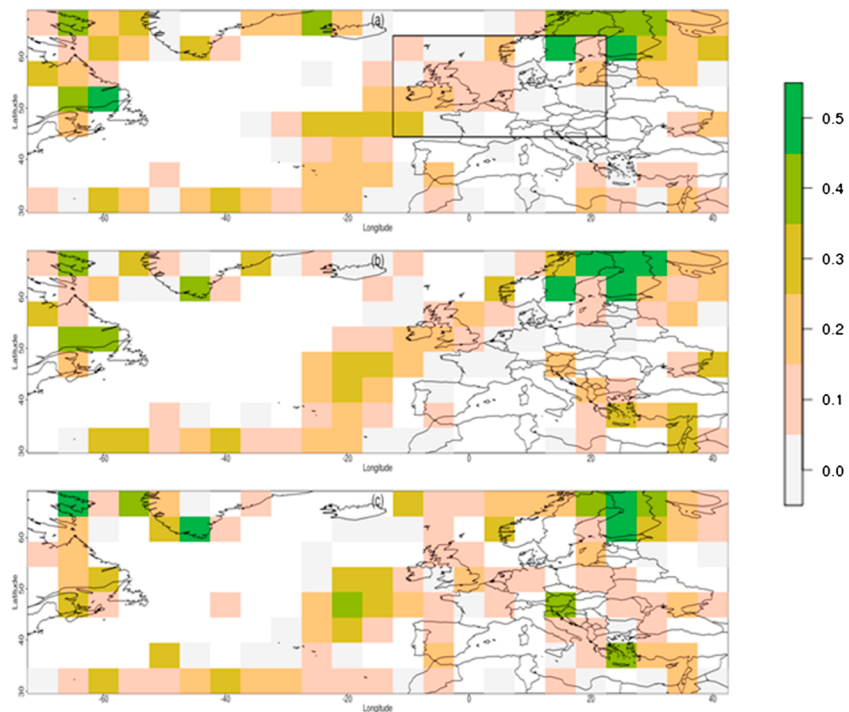
For the sake of completeness we present the results for model-specific results (utilizing System 4 eigenvectors in the supporting information). There the first EOF appears only occasionally as the most important MSLP variability pattern, whereas the PNA (second EOF), the WP (third EOF), and the East Pacific/North Pacific (eighth EOF) appear as the patterns explaining the most variability of extreme wind speeds in the North Atlantic area. The shifted importance of the different patterns appears to be related to the observed downstream displacement of the center of actions of the aforementioned MSLP variability patterns (Figure 2) as, for example, the second EOF (PNA), which explains variability in the European domain as well.

### 4.3. Predictability of High Wind Speeds in the ECMWF System 4

The estimated predictability (using mean PP) of  $I_{k,y,95}$  is presented in Figure 6 for all three evaluated percentiles. The figures show the average PP for every  $5 \times 5$  grid cell over the period from DJF 1982/1983 until DJF 2013/2014. For exceedances of all three percentile thresholds there seem to be three main pockets of large PP values: Over the Scandinavian region, over the Central-Eastern Atlantic, and over Eastern Canada/Newfoundland. Interestingly, two of these areas are toward the edge of the analyzed domain. The maximum PP across all three percentile exceedances is found in the Scandinavian region. This could be due to the fact that the prior uncertainty in these regions is relatively large as the climatology here exhibits a strong year-to-year variability. This indicates that the initialized runs provide useful information to narrow this spread for the upcoming winter season. The large climatological spread is possibly also the reason for the fact that there is no PP over the Atlantic within the classic storm track. The prior uncertainty in this region is relatively small in this region, and so the forecast does not provide any more useful information or is able to reduce the uncertainty.

The PP does generally not decrease with increasing percentile threshold. The overall pattern looks very similar in all three plots. Interestingly, certain areas feature even higher PP values for the higher percentiles (e.g., Northern Atlantic west of the British Isles or across Central Europe). Generally, System 4 appears to provide





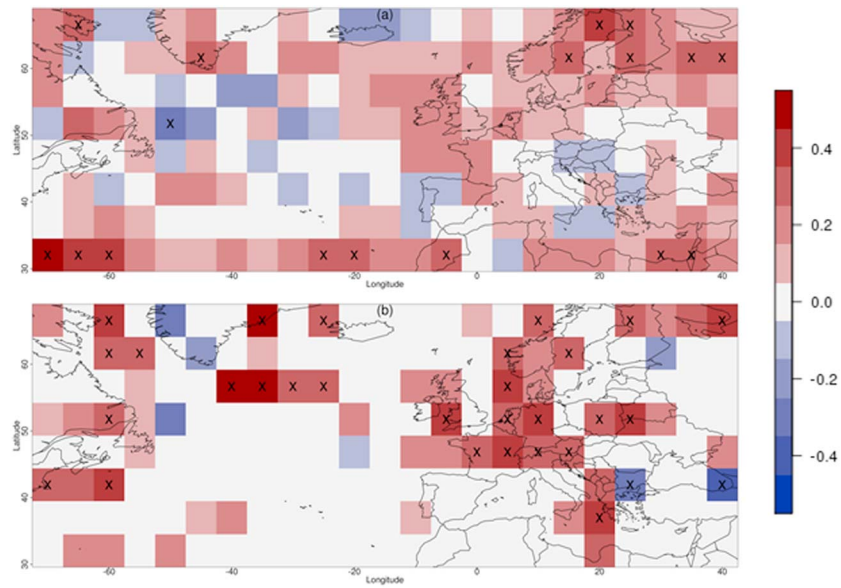
**Figure 6.** Mean PP of ECMWF System 4 for the DJF seasons from 1982/1983 to 2013/2014 for the extreme wind speeds (a)  $I_{k,y,90}$ , (b)  $I_{k,y,95}$ , and (c)  $I_{k,y,98}$ . The rectangular box marks the regions for which regional average time series are analyzed (see Figure 8). PP = predictive power; ECMWF = European Center for Medium-Range Weather Forecasts; DJF = December–January–February.

additional information to the climatology where the amount of windstorms associated with extreme winds has a larger interannual variability. This makes sense as the climatological distributions in these areas will be reasonably broad; thus, the prior uncertainty is large. From an end users' point of view this is very useful information as it seems that the forecast of high wind speeds exhibits a good degree of sharpness in areas where potential damage could occur from those winds.

#### 4.4. Evaluation of the Statistical Model in Comparison to System 4

The statistical model developed in section 3.5 is based on the nine leading EOFs of ERA-Interim mid-October to mid-November MSLP data. That way, it also attempts to predict winter storminess based on October to November initial conditions and ensures a fair level of comparison between the statistical model and the dynamical seasonal forecast product System 4. Figure 7 shows maps of the correlation between the respective model and ERA-Interim for the normalized sum of extreme winds  $I_{k,y,95}$  during DJF.

Compared to System 4, the statistical model undoubtedly shows more grid cells with a significant correlation (skill) for predicting the extreme wind speeds during the DJF season. The significant grid cells for System 4 tend to be at the edge of the studied domain, whereas the major part of grid cells with significant correlations for the statistical model appears over Central and Northern Europe featuring correlations up to 0.5. Interestingly, the correlation for System 4 is highest where there was also the highest magnitude of PP (e.g., the Scandinavian region in Figure 6), which confirms that the PP can be used as an alternative tool to examine the predictive skill of an ensemble forecast. A definite drawback to the statistical model is the fact that there is not necessarily a statistical model for every grid cell. In case none of the nine EOFs of the November MSLP turns out to be significantly related to local extreme wind speeds, no model can be determined. The maps look very similar for  $I_{k,y,90}$  and  $I_{k,y,98}$ , so the result is not sensitive to the wind speed threshold. Therefore, we choose not to show them at this point. Interestingly, many grid cells for which there is no statistical model exhibit some skill in the dynamical forecast. An interesting approach could be to combine the two forecast models. One possibility would be to check for skill in System 4. If none can be found the statistical model could be checked for a significant model in that particular grid cell. If skill can be found that grid

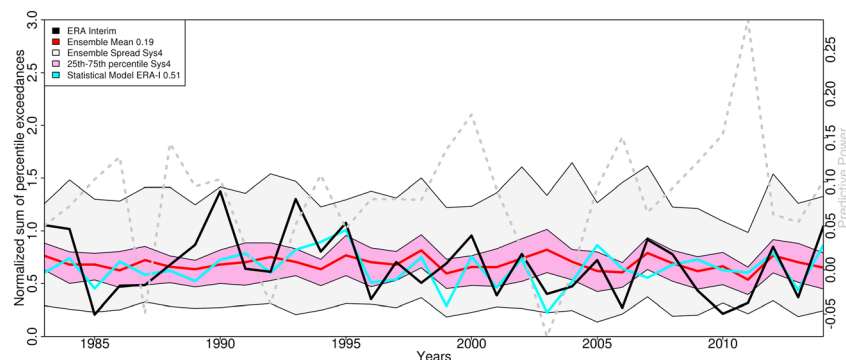


**Figure 7.** Correlation maps for  $I_{k,y,95}$  time series between (a) ERA-Interim and the ensemble mean of System 4 and (b) ERA-Interim and the statistical model based on November MSLP data. All significant grid cells at the 5% confidence level are marked with a black cross. ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; MSLP = mean sea level pressure.

cell could be used from the statistical model. Judging from Figure 7, this “hybrid” model would exhibit skill in larger areas than any individual model.

To include all of the previously discussed predictability scores and correlations into one plot, all of these items were averaged over the defined rectangular box (see Figure 6) including large parts of Central Europe and the British Isles, thus roughly the area that is most affected by European windstorms (e.g., Leckebusch et al., 2008).

The time series (Figure 8) of  $I_{k,y,95}$  in ERA-Interim nicely captures the interannual variability of extreme winds in Central Europe with peaks for the stormy seasons 1989/1990, 1992/1993, 1999/2000, 2006/2007, and 2013/2014 as these featured some of the most prominent European windstorms in the last 30 years. The



**Figure 8.** Time series of  $I_{k,y,95}$  spatially averaged for the European region as defined in Figure 6. The black line corresponds to ERA-Interim, the red line corresponds to the mean of System 4, the purple band depicts the inner 50th percentile of System 4, and the gray band depicts the range of the entire ensemble of System 4. The cyan line represents  $I_{k,y,95}$  as predicted by the statistical model. The dashed light gray line depicts the PP whose y-axis is plotted on the right-hand side of the plot. The correlations between the ensemble mean of System 4 and the statistical model with ERA-Interim are given in the legend after the respective label. The years on the x-axis refer to the years of the January and February of a particular DJF season; thus, 1990 refers to the winter 1989/1990, and so on. ERA = ERA = ECMWF Re-Analysis; EMCWF = European Center for Medium-Range Weather Forecasts; PP = predictive power; DJF = December–January–February.

minima are found for the winters 1984/85 and 2009/10. Both of these seasons also exhibited strong negative values of the NAO (see Figure 3). It seems that the ensemble spread of System 4 is approximately as big as the yearly variations of ERA-Interim; however, the mean of the ensemble shows very little variability over the entire time period. This small signal variability is a common feature when taking the mean of a large ensemble size as it is provided with the 51 members by System 4. Due to the setup of System 4, it is not possible to examine the interannual variability of each member as member 12 in year 1, for example, is not necessarily member 12 in year 2. Thus, the examination of each member individually would mix different ensemble members.

The correlation between the ensemble mean and ERA-Interim of 0.19 is not significant at the 95% significance level. Every winter season in the System 4 ensemble mean appears to be of a similar intensity, whereas the observations exhibit a strong year-to-year variability. Compared to the range of the entire ensemble, the inner 50th percentile appears sharper. This leads to the conclusion that many ensemble members end up in the same state and thus agree well to one another. The time mean of  $I_{k,y,95}$  over the entire time is very similar for System 4 and ERA-Interim. This implies that the mean of the ensemble forecast manages to get the mean extremeness of wind speeds correct over the 32 years; however, it fails to predict the interannual variability.

The interannual variability in the statistical model is considerably larger. In addition, it correctly predicts some of the peaks (e.g., 1999/2000 or 2011/2012) although it fails to predict the prominent winter of 1989/1990. The statistical model occasionally predicts more extreme seasons than ERA-Interim (2001/2002 or 2009/2010). The correlation of 0.51 is significant at the 95% significance level.

Generally, it seems that unusually low extreme wind seasons like 1984/1985 and 2009/2010 are harder to predict for both the dynamical and the statistical model. The PP is negatively correlated ( $-0.5$ ) with the spread of the ensembles that is in line with the assumption that less spread/uncertainty in the forecast is related to a higher predictability (information gain) of the extreme wind speeds. Despite being less uncertain, the ensemble members appear to agree more with each other than with the actual observation. This compares to an underconfident forecast as discussed in Eade et al. (2014). In other words, there are barely any seasons in which the ensemble mean forecast can make a prediction of the actual intensity (peak intensity) of the season. In terms of information gain, however, the seasonal forecast still provides additional information to the climatology for some years, in particular for the winters 1999/2000, 2006/2007, and 2010/2011 (see the peaks of the PP curve in Figure 8).

## 5. Summary and Discussion

This study utilized the ECMWF System 4 and ERA-Interim to advance the understanding of the large-scale drivers associated with extreme winds (i.e., associated with European winter windstorms) and their predictability.

The NAO (first EOF) pattern in the seasonal forecast seems to be shifted downstream so that its center of action is relocated toward the east. This downstream displacement of variability modes is also observed for the second EOF (corresponding to PNA in the reanalysis) and the third EOF (WP), which are extended into the Atlantic.

The correlation of the two EOF-based NAO time series between the reanalysis and the projected System 4 index is around 0.49, whereas it is only 0.32 for the grid-cell-based index. The lower correlation for the latter could potentially be explained by the shifted location of the EOFs in System 4, similar to that in Ulbrich and Christoph (1999) who argued that the rigid location for the grid-cell-based NAO index could be inaccurate for some future general circulation model predictions (ECHAM4 + OPYC3 in their case) as the center of action of the NAO is displaced downstream. A similar displacement is observed for the seasonal forecast. The correlation skill of the NAO in System 4 is in good agreement with the findings of O'Reilly et al. (2017). They computed a correlation of 0.31 for the NAO index derived from 500-hPa geopotential height data using the (atmosphere-only) seasonal experiments for the entire twentieth century (Weisheimer et al., 2017) and of around 0.40 for the 31 years between 1980 and 2010, thus close to our estimated correlation of 0.49. The difference might be partly due to the slightly different time period used in their study.

The question about the predictability or uncertainty of extreme wind speeds in System 4 was approached with the statistical entropy and the PP (Schneider & Griffies, 1999; Tang et al., 2007) in particular. Evidently, there is more information in the seasonal forecast (larger PP value) in areas where the variability of the

climatology is fairly large. Thus, the seasonal forecast is able to reduce the uncertainty of the forecast distribution compared to the climatological forecast. This applies especially for some regions in the Central East Atlantic, Eastern Canada, and northern Scandinavia. The variability in the climatology of extreme winds within the prominent storm track is comparatively low so that the seasonal forecast cannot provide much additional information for these areas. Overall, the PP value proves to be a valid measure of uncertainty quantification. Fernández-González et al. (2017) investigated the uncertainty of wind speeds for the ECMWF ensemble prediction system (EPS) operational weather forecasting ensemble with a predictability index that evaluates the average interquartile range for a 30-day period against the current forecast. In a way, this is similar to the PP as it estimates in how far the uncertainty of the climatology can be reduced by a forecast, thus whether or not the predictability is better than “usually.” Overall, the conclusions they draw from their analysis of the ECMWF EPS are in line (see below) with System 4 although the forecasts are on different time scales and in their case on a much smaller spatial scale.

The map of the physical drivers of extreme wind speeds for System 4 looks fairly different to the reanalysis. The NAO acts as the dominant driver in the North Atlantic; however, it does not seem to explain much variability in the Central Atlantic. The sixth EOF (SCA) is identified as the main driver for extreme winds over Central Europe. When the model-specific eigenvectors are used to produce the time series, the differences between ERA-Interim and System 4 are even more striking. It has to be kept in mind here that we used the EOF patterns of ERA-Interim to create the EOF time series for System 4. By doing that, we could ensure a fairer comparison with regard to the selection of drivers; however, the question remains as to why the EOF patterns look fairly different between System 4 and ERA-Interim. Physically, a system generating extreme wind speeds (i.e., an extreme cyclone in the North Atlantic) should be the same within both System 4 and ERA-Interim. The actual EOF patterns shown in Figures 1 and 2 should in fact affect the probability of occurrence of these large-scale systems. We therefore cannot make any assumptions on how often the large-scale conditions generating windstorms in ERA-Interim actually occur in System 4 and how much variability of extreme wind speeds they can explain.

Even though the NAO in ERA-Interim and System 4 are sufficiently well correlated as shown in this study and also by Kim et al. (2012) or Weisheimer et al. (2017), the System 4 internal NAO fails to act as the most dominant driver associated with the interannual wind speed variability in Central Europe. This would suggest that the skillful prediction of the NAO, which is certainly possible (as shown, e.g., with GloSea5 by Scaife et al., 2014), may be used to deduce the winter storminess in the real world building on empirical relationships between NAO and storminess. However, it seems that the NAO and the winter storminess in the model itself are not as related as in the real world.

On the one hand, this would mean that extreme wind speeds in the European region are physically associated with different variability patterns even if the large-scale wind speed variability pattern looks correct (“being right for the wrong reasons”). On the other hand, this also means that by improving the model-internal MSLP variability, the skill of the seasonal forecast could potentially be substantially increased. Of course, this would not be constrained to the skill of extreme winds but may also affect the skill of various other atmospheric variables.

The spatially averaged time series revealed a major inconsistency of System 4 in comparison to the ERA-Interim reanalysis: Whereas the time mean over the 32-year period of System 4 and ERA-Interim agree well, the year-to-year variability of extreme winds is clearly underestimated by System 4. This lack of variability is a common feature of an ensemble mean. An examination of the variability of the individual members, however, would be fallacious as the members are not consistent for different initializations in time. The spread of the interquartile range of System 4 appears fairly sharp, implying that at least half of the ensemble members predict wind speeds relatively close to each other. This is in accordance with Fernández-González et al. (2017) who determined the interquartile range as the most balanced uncertainty quantification for wind speeds over Spain in the ECMWF EPS forecasting system. They also note that the ensemble mean should be accompanied by some measure of uncertainty. Based on Figure 8, we can confirm this claim.

While the System 4 prediction skill for extreme winds is relatively low, possible avenues for improvements may be to use multisystem approaches. For example, Athanasiadis et al. (2017) could show that a multisystem of seasonal forecasts consisting of GloSea5, the National Center for Environmental Prediction Climate Forecast System version 2 (Saha et al., 2014), and the Euro-Mediterranean Center on Climate Change



(CMCC) Seasonal Prediction System 1.5 (Materia et al., 2014) has unprecedented high predictive (correlation) skill for the NAO and Arctic Oscillation of 0.85 for the short period of 1997–2014. They follow the arguments by DelSole et al. (2014) who found that the enhanced predictive skill of a multisystem of forecasts exceeds the expected increase due to larger ensemble sizes alone. DelSole et al. (2014) argue that the increase in predictive skill is more consistent with the addition of new signals from the different forecast systems. Thus, one way to possibly improve the forecasting skill would be to incorporate more than one seasonal forecast product for future research.

Befort et al. (2018) showed significant skill in their event-based study analyzing the predictability of European windstorm occurrence in seasonal forecasts. An explanation for the higher skill in their study could be that the interannual variability in the frequency of windstorms in the North Atlantic is less affected by the variability of extreme wind speeds at a particular grid cell. The tracking algorithm used to identify windstorms is based on an exceedance of a local percentile. The magnitude of the exceedance of the percentile, however, is irrelevant for the identification of an event, whereas it is relevant for our  $I_{k,y,9X}$  value. Thus, the degree of exceedance adds an extra dimension of uncertainty to the skill assessment.

It has to be noted, also, that their analysis only comprises the years 1992–2011 and utilizes wind speeds in 925 hPa, whereas we use 10-m wind speeds over the period 1982–2013. We did some brief analysis only taking into account this shorter time span and found correlations of similar magnitude and pattern to what they had found in their paper, for example, off the coast of France and the United Kingdom in particular (not shown). Whereas their results appear significant, our map of correlations for the period from 1992–2011 still shows very little statistical significance. The finding of lower skill for the 1982–2013 period is also in line with their findings.

The two main conclusions of our study are as follows:

1. The EOF patterns of System 4 look considerably different compared to the reanalysis. Even when projecting the MSLP anomalies of System 4 onto ERA-Interim loading patterns, the large-scale drivers correlated with extreme winds in System 4 are fairly different to what is observed in the reanalysis. This could be one of the reasons for the low skill of the ensemble mean with regard to the observed value of  $I_{k,y,9X}$ . The low skill is in accordance with Della-Marta et al. (2010) who found no skill for the 95th percentile of wind speeds in the predecessor of System 4 for months past the first lead month (lead months 2–4). The open question here remains as to why the coupled model that runs freely after initialization produces such different patterns of atmospheric variability compared to the reanalysis that is constraint by observations. Particularly, the generating systems of extreme winds in DJF (extreme windstorms) should be the same in both System 4 and ERA-Interim.
2. A cross-validated multilinear regression model using large-scale MSLP patterns from mid-October to mid-November can provide significant skill for the upcoming DJF season with regard to extreme wind speeds. Thus, October and November MSLP data contain valuable information regarding the upcoming DJF storminess for vulnerable regions, pointing to potential predictability in the climate system that may not be captured by the numerical model in System 4. The statistical model could potentially be further improved by incorporating longer time series, for example, by using ECMWF ERA-20C (Poli et al., 2016).

For future research it would be interesting to investigate whether there is a temporal progression in the spatial shift in System 4 compared to ERA-Interim from initialization to increasing lead times, for example, carry out an EOF analysis for different lead times. Additionally, the identification of drivers could be applied to other variables, especially to those that have already been validated to feature higher skill (e.g., winter temperature; Kim et al., 2012, or Ogutu et al., 2017) than the erratic wind speeds. The statistical model developed in this study provides a good indication about the potentially predictable storminess of the upcoming winter season and could provide a useful tool for the impact/actuarial community.

#### Acknowledgments

M. A. Walz has been supported by a NERC CENTA DTP PhD award kindly funded by Research Council UK. For his visit to the University of New South Wales he was funded by the Universitas 21 program as well as by the Legacies Fund of the Royal Meteorological Society. M. G. D. received funding from Australian Research Council grant DE150100456. The authors thank the ECMWF for providing the Seasonal System 4 data and also ERA-Interim reanalysis data. The authors would also like to thank four anonymous reviewers that have helped to improve the quality of the paper. All the data/results of this paper were created from ERA-Interim (Dee et al., 2011) and ECMWF System 4 (Molteni et al., 2011) data, which are freely available for academic use through the ECMWF server via <http://apps.ecmwf.int/archive-catalogue/> and <https://www.ecmwf.int/en/forecasts/accessing-forecasts/order-historical-datasets> (MARS access), respectively.

#### References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Ambaum, M. H., Hoskins, B. J., & Stephenson, D. B. (2001). Arctic Oscillation or North Atlantic Oscillation? *Journal of Climate*, 14(16), 3495–3507. [https://doi.org/10.1175/1520-0442\(2001\)014<3495:AOONAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3495:AOONAO>2.0.CO;2)
- Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F., Molteni, F., et al. (2007). Development of the ECMWF seasonal forecast system 3. *ECMWF Technical Memoranda*, 503.

- Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Materia, S., Sanna, A., et al. (2017). A multisystem view of wintertime NAO seasonal predictions. *Journal of Climate*, 30(4), 1461–1475. <https://doi.org/10.1175/JCLI-D-16-0153.1>
- Befort, D. J., Wild, S., Knight, J. R., Lockwood, J. F., Thornton, H. E., Hermanson, L., et al. (2018). Seasonal forecast skill for extra-tropical cyclones and windstorms. *Quarterly Journal of the Royal Meteorological Society*. <https://doi.org/10.1002/qj.3406>
- Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>
- Della-Marta, P. M., Liniger, M. A., Appenzeller, C., Bresch, D. N., Köllner-Heck, P., & Muccione, V. (2010). Improved estimates of the European winter windstorm climate and the risk of reinsurance loss using climate model data. *Journal of Applied Meteorology and Climatology*, 49(10), 2092–2120. <https://doi.org/10.1175/2010JAMC2133.1>
- DelSole, T. (2004). Predictability and information theory. Part I: Measures of predictability. *Journal of the Atmospheric Sciences*, 61(20), 2425–2440. [https://doi.org/10.1175/1520-0469\(2004\)061<2425:PAITP>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<2425:PAITP>2.0.CO;2)
- DelSole, T., Nattala, J., & Tippett, M. K. (2014). Skill improvement from increased ensemble size and model diversity. *Geophysical Research Letters*, 41, 7331–7342. <https://doi.org/10.1002/2014GL060133>
- Doblas-Reyes, F. J., García-Serrano, J., Lienert, F., Biescas, A. P., & Rodrigues, L. R. (2013). Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdisciplinary Reviews: Climate Change*, 4(4), 245–268.
- Donat, M. G., Leckebusch, G. C., Pinto, J. G., & Ulbrich, U. (2010). Examination of wind storms over Central Europe with respect to circulation weather types and NAO phases. *International Journal of Climatology*, 30(9), 1289–1300.
- Eade, R., Smith, D., Scaife, A., Wallace, E., Dunstone, N., Hermanson, L., & Robinson, N. (2014). Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophysical Research Letters*, 41, 5620–5628. <https://doi.org/10.1002/2014GL061146>
- Easterling, D. R., Evans, J., Groisman, P. Y., Karl, T., Kunkel, K. E., & Ambenje, P. (2000). Observed variability and trends in extreme climate events: A brief review. *Bulletin of the American Meteorological Society*, 81(3), 417–425. [https://doi.org/10.1175/1520-0477\(2000\)081<0417:OVATIE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2)
- European Environment Agency (2011). *Mapping the impacts of natural hazards and technological accidents in Europe: An overview of the last decade*. Copenhagen: Publications Office of the European Union.
- Fernández-González, S., Martín, M. L., Merino, A., Sánchez, J. L., & Valero, F. (2017). Uncertainty quantification and predictability of wind speed over the Iberian Peninsula. *Journal of Geophysical Research: Atmospheres*, 122, 3877–3890. <https://doi.org/10.1002/2017JD026533>
- Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., et al. (2013). A verification framework for interannual-to-decadal predictions experiments. *Climate Dynamics*, 40(1–2), 245–272. <https://doi.org/10.1007/s00382-012-1481-2>
- Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269(5224), 676–679. <https://doi.org/10.1126/science.269.5224.676>
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620–630. <https://doi.org/10.1103/PhysRev.106.620>
- Jolliffe, I. T. (1986). *Principal component analysis and factor analysis*. *Principal component analysis*, (pp. 115–128). Springer: Springer.
- Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dynamics*, 39(12), 2957–2973. <https://doi.org/10.1007/s00382-012-1364-6>
- Knight, J. R., Folland, C. K., & Scaife, A. A. (2006). Climate impacts of the Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, 33, L17706. <https://doi.org/10.1029/2006GL026242>
- Leckebusch, G. C., Renggli, D., & Ulbrich, U. (2008). Development and application of an objective storm severity measure for the Northeast Atlantic region. *Meteorologische Zeitschrift*, 17(5), 575–587. <https://doi.org/10.1127/0941-2948/2008/0323>
- Lorenz, E. N. (1963). Section of planetary sciences: The predictability of hydrodynamic flow. *Transactions of the New York Academy of Sciences*, 25(4 Series II), 409–432. <https://doi.org/10.1111/j.2164-0947.1963.tb01464.x>
- Lowe, R., García-Díez, M., Ballester, J., Creswick, J., Robine, J. M., Herrmann, F. R., & Rodó, X. (2016). Evaluation of an early-warning system for heat wave-related mortality in Europe: implications for sub-seasonal to seasonal forecasting and climate services. *International Journal of Environmental Research and Public Health*, 13(2), 206.
- MacLachlan, C., Arribas, A., Peterson, K., Maidens, A., Fereday, D., Scaife, A., et al. (2015). Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quarterly Journal of the Royal Meteorological Society*, 141(689), 1072–1084. <https://doi.org/10.1002/qj.2396>
- MacLeod, D., Torralba, V., Davis, M., & Doblas-Reyes, F. (2018). Transforming climate model output to forecasts of wind power production: How much resolution is enough? *Meteorological Applications*, 25(1), 1–10. <https://doi.org/10.1002/met.1660>
- Madec, G. (2008). *NEMO reference manual, ocean dynamics component: NEMO-OPA. Preliminary version*. Paris: Note du Pôle de modélisation 27. Institut Pierre-Simon Laplace.
- Mailier, P. J., Stephenson, D. B., Ferro, C. A., & Hodges, K. I. (2006). Serial clustering of extratropical cyclones. *Monthly Weather Review*, 134(8), 2224–2240. <https://doi.org/10.1175/MWR3160.1>
- Materia, S., Borrelli, A., Bellucci, A., Alessandri, A., Di Pietro, P., Athanasiadis, P., et al. (2014). Impact of atmosphere and land surface initial conditions on seasonal forecasts of global surface temperature. *Journal of Climate*, 27(24), 9253–9271. <https://doi.org/10.1175/JCLI-D-14-00163.1>
- Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., et al. (2011). The new ECMWF seasonal forecast system (System 4). Technical Memorandum, 656.
- Nissen, K. M., Ulbrich, U., Leckebusch, G. C., & Kuhnel, I. (2014). Decadal windstorm activity in the North Atlantic-European sector and its relationship to the meridional overturning circulation in an ensemble of simulations with a coupled climate model. *Climate Dynamics*, 43(5–6), 1545–1555. <https://doi.org/10.1007/s00382-013-1975-6>
- Ogutu, G. E., Franssen, W. H., Supit, I., Omondi, P., & Hutjes, R. W. (2017). Skill of ECMWF System-4 ensemble seasonal climate forecasts for East Africa. *International Journal of Climatology*, 37(5), 2734–2756. <https://doi.org/10.1002/joc.4876>
- O'Reilly, C. H., Heatley, J., MacLeod, D., Weisheimer, A., Palmer, T. N., Schaller, N., & Woollings, T. (2017). Variability in seasonal forecast skill of Northern Hemisphere winters over the twentieth century. *Geophysical Research Letters*, 44, 5729–5738. <https://doi.org/10.1002/2017GL073736>
- Palmer, T., Doblas-Reyes, F., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., et al. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85(6), 853–872. <https://doi.org/10.1175/BAMS-85-6-853>
- Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., & Ulbrich, U. (2009). Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate Dynamics*, 32(5), 711–737. <https://doi.org/10.1007/s00382-008-0396-4>
- Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11), 4083–4097. <https://doi.org/10.1175/JCLI-D-15-0556.1>

- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, 27(6), 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Scaife, A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, 41, 2514–2519. <https://doi.org/10.1002/2014GL059637>
- Schneider, T., & Griffies, S. M. (1999). A conceptual framework for predictability studies. *Journal of Climate*, 12(10), 3133–3155. [https://doi.org/10.1175/1520-0442\(1999\)012<3133:ACFFPS>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<3133:ACFFPS>2.0.CO;2)
- Shannon, C. E. (1948). A note on the concept of entropy. *Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell Labs Technical Journal*, 30(1), 50–64. <https://doi.org/10.1002/j.1538-7305.1951.tb01366.x>
- Tang, Y., Lin, H., Derome, J., & Tippett, M. K. (2007). A predictability measure applied to seasonal predictions of the Arctic Oscillation. *Journal of Climate*, 20(18), 4733–4750. <https://doi.org/10.1175/JCLI4276.1>
- Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I., & Davis, M. (2017). Seasonal climate prediction: A new source of information for the management of wind energy resources. *Journal of Applied Meteorology and Climatology*, 56(5), 1231–1247. <https://doi.org/10.1175/JAMC-D-16-0204.1>
- Ulbrich, U., & Christoph, M. (1999). A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing. *Climate Dynamics*, 15(7), 551–559. <https://doi.org/10.1007/s003820050299>
- Walz, M. A., Befort, D. J., Kirchner-Bossi, N. O., Ulbrich, U., & Leckebusch, G. C. (2018). Modelling serial clustering and inter-annual variability of European winter windstorms based on largescale drivers. *International Journal of Climatology*, 38(7), 3044–3057. <https://doi.org/10.1002/joc.5481>
- Weisheimer, A., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, 143(703), 917–926. <https://doi.org/10.1002/qj.2976>
- Woollings, T., Franzke, C., Hodson, D. L. R., Dong, B., Barnes, E. A., Raible, C. C., & Pinto, J. G. (2015). Contrasting interannual and multidecadal NAO variability. *Climate Dynamics*, 45(1–2), 539–556. <https://doi.org/10.1007/s00382-014-2237-y>
- Woollings, T., Gregory, J. M., Pinto, J. G., Meyers, M., & Brayshaw, D. J. (2012). Response of the North Atlantic storm track to climate change shaped by ocean-atmosphere coupling. *Nature Geoscience*, 5(5), 313–317. <https://doi.org/10.1038/ngeo1438>